



PCT10

## RAW SEQUENCE LISTING

DATE: 08/06/2002

PATENT APPLICATION: US/10/031,695

TIME: 10:35:30

Input Set : A:\031695s1

Output Set: N:\CRF3\08062002\J031695.raw

4 <110> APPLICANT: Hauer, Bernhard  
 5 Pleiss, Jurgen  
 6 Schwaneberg, Ulrich  
 7 Schmitt, Jutta  
 9 <120> TITLE OF INVENTION: Modified cytochrome P450 monooxygenases  
 11 <130> FILE REFERENCE: M/40434  
 W--> 12 <140> CURRENT APPLICATION NUMBER: US 10/031,695  
 C--> 13 <141> CURRENT FILING DATE: 2002-05-06  
 13 <150> PRIOR APPLICATION NUMBER: PCT/EP00/07252  
 14 <151> PRIOR FILING DATE: 2000-07-27  
 16 <160> NUMBER OF SEQ ID NOS: 20  
 17 <170> SOFTWARE: PatentIn Ver. 2.1  
 20 <210> SEQ ID NO: 1  
 21 <211> LENGTH: 3150  
 22 <212> TYPE: DNA  
 23 <213> ORGANISM: Bacillus megaterium  
 25 <220> FEATURE:  
 26 <221> NAME/KEY: CDS  
 27 <222> LOCATION: (4)..(3150)  
 29 <400> SEQUENCE: 1  
 31 atg aca att aaa gaa atg cct cag cca aaa acg ttt gga gag ctt aaa 48  
 32 Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys  
 33 1 5 10 15  
 35 aat tta ccg tta tta aac aca gat aaa ccg gtt caa gct ttg atg aaa 96  
 36 Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys  
 37 20 25 30  
 39 att gcg gat gaa tta gga gaa atc ttt aaa ttc gag gcg cct ggt cgt 144  
 40 Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Arg  
 41 35 40 45  
 43 gta acg cgc tac tta tca agt cag cgt cta att aaa gaa gca tgc gat 192  
 44 Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp  
 45 50 55 60  
 47 gaa tca cgc ttt gat aaa aac tta agt caa gcg ctt aaa ttt gta cgt 240  
 48 Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Val Arg  
 49 65 70 75  
 51 gat ttt gca gga gac ggg tta ttt aca agc tgg acg cat gaa aaa aat 288  
 52 Asp Phe Ala Gly Asp Gly Leu Phe Thr Ser Trp Thr His Glu Lys Asn  
 53 80 85 90 95  
 55 tgg aaa aaa gcg cat aat atc tta ctt cca agc ttc agt cag cag gca 336  
 56 Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala  
 57 100 105 110  
 59 atg aaa ggc tat cat gcg atg atg gtc gat atc gcc gtg cag ctt gtt 384  
 60 Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val

ENTERED

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/10/031,695

DATE: 08/06/2002

TIME: 10:35:30

Input Set : A:\031695s1

Output Set: N:\CRF3\08062002\J031695.raw

61	115	120	125	
63	caa aag tgg gag cgt cta aat gca gat gag cat att gaa gta ccg gaa	432		
64	Gln Lys Trp Glu Arg Leu Asn Ala Asp Glu His Ile Glu Val Pro Glu			
65	130	135	140	
67	gac atg aca cgt tta acg ctt gat aca att ggt ctt tgc ggc ttt aac	480		
68	Asp Met Thr Arg Leu Thr Leu Asp Thr Ile Gly Leu Cys Gly Phe Asn			
69	145	150	155	
71	tat cgc ttt aac agc ttt tac cga gat cag cct cat cca ttt att aca	528		
72	Tyr Arg Phe Asn Ser Phe Tyr Arg Asp Gln Pro His Pro Phe Ile Thr			
73	160	165	170	175
75	agt atg gtc cgt gca ctg gat gaa gca atg aac aag ctg cag cga gca	576		
76	Ser Met Val Arg Ala Leu Asp Glu Ala Met Asn Lys Leu Gln Arg Ala			
77	180	185	190	
79	aat cca gac gac cca gct tat gat gaa aac aag cgc cag ttt caa gaa	624		
80	Asn Pro Asp Asp Pro Ala Tyr Asp Glu Asn Lys Arg Gln Phe Gln Glu			
81	195	200	205	
83	gat atc aag gtg atg aac gac cta gta gat aaa att att gca gat cgc	672		
84	Asp Ile Lys Val Met Asn Asp Leu Val Asp Lys Ile Ile Ala Asp Arg			
85	210	215	220	
87	aaa gca agc ggt gaa caa agc gat gat tta tta acg cat atg cta aac	720		
88	Lys Ala Ser Gly Glu Gln Ser Asp Asp Leu Leu Thr His Met Leu Asn			
89	225	230	235	
91	gga aaa gat cca gaa acg ggt gag ccg ctt gat gac gag aac att cgc	768		
92	Gly Lys Asp Pro Glu Thr Gly Glu Pro Leu Asp Asp Glu Asn Ile Arg			
93	240	245	250	255
95	tat caa att att aca ttc tta att gcg gga cac gaa aca aca agt ggt	816		
96	Tyr Gln Ile Ile Thr Phe Leu Ile Ala Gly His Glu Thr Thr Ser Gly			
97	260	265	270	
99	ctt tta tca ttt gcg ctg tat ttc tta gtg aaa aat cca cat gta tta	864		
100	Leu Leu Ser Phe Ala Leu Tyr Phe Leu Val Lys Asn Pro His Val Leu			
101	275	280	285	
103	caa aaa gca gca gaa gaa gca gca cga gtt cta gta gat cct gtt cca	912		
104	Gln Lys Ala Ala Glu Glu Ala Ala Arg Val Leu Val Asp Pro Val Pro			
105	290	295	300	
107	agc tac aaa caa gtc aaa cag ctt aaa tat gtc ggc atg gtc tta aac	960		
108	Ser Tyr Lys Gln Val Lys Gln Leu Lys Tyr Val Gly Met Val Leu Asn			
109	305	310	315	
111	gaa gcg ctg cgc tta tgg cca act gct cct gcg ttt tcc cta tat gca	1008		
112	Glu Ala Leu Arg Leu Trp Pro Thr Ala Pro Ala Phe Ser Leu Tyr Ala			
113	320	325	330	335
115	aaa gaa gat acg gtg ctt gga gga gaa tat cct tta gaa aaa ggc gac	1056		
116	Lys Glu Asp Thr Val Leu Gly Gly Glu Tyr Pro Leu Glu Lys Gly Asp			
117	340	345	350	
119	gaa cta atg gtt ctg att cct cag ctt cac cgt gat aaa aca att tgg	1104		
120	Glu Leu Met Val Leu Ile Pro Gln Leu His Arg Asp Lys Thr Ile Trp			
121	355	360	365	
123	gga gac gat gtg gaa gag ttc cgt cca gag cgt ttt gaa aat cca agt	1152		
124	Gly Asp Asp Val Glu Glu Phe Arg Pro Glu Arg Phe Glu Asn Pro Ser			
125	370	375	380	

## RAW SEQUENCE LISTING

DATE: 08/06/2002

PATENT APPLICATION: US/10/031,695

TIME: 10:35:30

Input Set : A:\031695s1

Output Set: N:\CRF3\08062002\J031695.raw

127	gcg	att	ccg	cag	cat	gcg	ttt	aaa	ccg	ttt	gga	aac	ggt	cag	cgt	gcg	1200
128	Ala	Ile	Pro	Gln	His	Ala	Phe	Lys	Pro	Phe	Gly	Asn	Gly	Gln	Arg	Ala	
129	385						390					395					
131	tgt	atc	ggt	cag	cag	ttc	gct	ctt	cat	gaa	gca	acg	ctg	gta	ctt	ggt	1248
132	Cys	Ile	Gly	Gln	Gln	Phe	Ala	Leu	His	Glu	Ala	Thr	Leu	Val	Leu	Gly	
133	400					405					410					415	
135	atg	atg	cta	aaa	cac	ttt	gac	ttt	gaa	gat	cat	aca	aac	tac	gag	ctg	1296
136	Met	Met	Leu	Lys	His	Phe	Asp	Phe	Glu	Asp	His	Thr	Asn	Tyr	Glu	Leu	
137					420						425					430	
139	gat	att	aaa	gaa	act	tta	acg	tta	aaa	cct	gaa	ggc	ttt	gtg	gta	aaa	1344
140	Asp	Ile	Lys	Glu	Thr	Leu	Thr	Leu	Lys	Pro	Glu	Gly	Phe	Val	Val	Lys	
141					435					440				445			
143	gca	aaa	tcg	aaa	aaa	att	ccg	ctt	ggc	ggt	att	cct	tca	cct	agc	act	1392
144	Ala	Lys	Ser	Lys	Lys	Ile	Pro	Leu	Gly	Gly	Ile	Pro	Ser	Pro	Ser	Thr	
145			450					455					460				
147	gaa	cag	tct	gct	aaa	aaa	gta	cgc	aaa	aag	gca	gaa	aac	gct	cat	aat	1440
148	Glu	Gln	Ser	Ala	Lys	Lys	Val	Arg	Lys	Lys	Ala	Glu	Asn	Ala	His	Asn	
149			465				470					475					
151	acg	ccg	ctg	ctt	gtg	cta	tac	ggt	tca	aat	atg	gga	aca	gct	gaa	gga	1488
152	Thr	Pro	Leu	Leu	Val	Leu	Tyr	Gly	Ser	Asn	Met	Gly	Thr	Ala	Glu	Gly	
153	480					485					490					495	
155	acg	gcg	cgt	gat	tta	gca	gat	att	gca	atg	agc	aaa	gga	ttt	gca	ccg	1536
156	Thr	Ala	Arg	Asp	Leu	Ala	Asp	Ile	Ala	Met	Ser	Lys	Gly	Phe	Ala	Pro	
157					500					505					510		
159	cag	gtc	gca	acg	ctt	gat	tca	cac	gcc	gga	aat	ctt	ccg	cgc	gaa	gga	1584
160	Gln	Val	Ala	Thr	Leu	Asp	Ser	His	Ala	Gly	Asn	Leu	Pro	Arg	Glu	Gly	
161					515				520					525			
163	gct	gta	tta	att	gta	acg	gcg	tct	tat	aac	ggt	cat	ccg	cct	gat	aac	1632
164	Ala	Val	Leu	Ile	Val	Thr	Ala	Ser	Tyr	Asn	Gly	His	Pro	Pro	Asp	Asn	
165			530					535					540				
167	gca	aag	caa	ttt	gtc	gac	tgg	tta	gac	caa	gcg	tct	gct	gat	gaa	gta	1680
168	Ala	Lys	Gln	Phe	Val	Asp	Trp	Leu	Asp	Gln	Ala	Ser	Ala	Asp	Glu	Val	
169			545				550				555						
171	aaa	ggc	gtt	cgc	tac	tcc	gta	ttt	gga	tgc	ggc	gat	aaa	aac	tgg	gct	1728
172	Lys	Gly	Val	Arg	Tyr	Ser	Val	Phe	Gly	Cys	Gly	Asp	Lys	Asn	Trp	Ala	
173	560					565					570					575	
175	act	acg	tat	caa	aaa	gtg	cct	gct	ttt	atc	gat	gaa	acg	ctt	gcc	gct	1776
176	Thr	Thr	Tyr	Gln	Lys	Val	Pro	Ala	Phe	Ile	Asp	Glu	Thr	Leu	Ala	Ala	
177					580					585					590		
179	aaa	ggg	gca	gaa	aac	atc	gct	gac	cgc	ggt	gaa	gca	gat	gca	agc	gac	1824
180	Lys	Gly	Ala	Glu	Asn	Ile	Ala	Asp	Arg	Gly	Glu	Ala	Asp	Ala	Ser	Asp	
181				595					600					605			
183	gac	ttt	gaa	ggc	aca	tat	gaa	gaa	tgg	cgt	gaa	cat	atg	tgg	agt	gac	1872
184	Asp	Phe	Glu	Gly	Thr	Tyr	Glu	Glu	Trp	Arg	Glu	His	Met	Trp	Ser	Asp	
185			610					615					620				
187	gta	gca	gcc	tac	ttt	aac	ctc	gac	att	gaa	aac	agt	gaa	gat	aat	aaa	1920
188	Val	Ala	Ala	Tyr	Phe	Asn	Leu	Asp	Ile	Glu	Asn	Ser	Glu	Asp	Asn	Lys	
189			625				630				635						
191	tct	act	ctt	tca	ctt	caa	ttt	gtc	gac	agc	gcc	gcg	gat	atg	ccg	ctt	1968

## RAW SEQUENCE LISTING

DATE: 08/06/2002

PATENT APPLICATION: US/10/031,695

TIME: 10:35:30

Input Set : A:\031695s1

Output Set: N:\CRF3\08062002\J031695.raw

192	Ser	Thr	Leu	Ser	Leu	Gln	Phe	Val	Asp	Ser	Ala	Ala	Asp	Met	Pro	Leu	
193	640					645					650					655	
195	gcg	aaa	atg	cac	ggt	gcg	ttt	tca	acg	aac	gtc	gta	gca	agc	aaa	gaa	2016
196	Ala	Lys	Met	His	Gly	Ala	Phe	Ser	Thr	Asn	Val	Val	Ala	Ser	Lys	Glu	
197					660					665					670		
199	ctt	caa	cag	cca	ggc	agt	gca	cga	agc	acg	cga	cat	ctt	gaa	att	gaa	2064
200	Leu	Gln	Gln	Pro	Gly	Ser	Ala	Arg	Ser	Thr	Arg	His	Leu	Glu	Ile	Glu	
201					675					680					685		
203	ctt	cca	aaa	gaa	gct	tct	tat	caa	gaa	gga	gat	cat	tta	ggt	gtt	att	2112
204	Leu	Pro	Lys	Glu	Ala	Ser	Tyr	Gln	Glu	Gly	Asp	His	Leu	Gly	Val	Ile	
205					690					695					700		
207	cct	cgc	aac	tat	gaa	gga	ata	gta	aac	cgt	gta	aca	gca	agg	ttc	ggc	2160
208	Pro	Arg	Asn	Tyr	Glu	Gly	Ile	Val	Asn	Arg	Val	Thr	Ala	Arg	Phe	Gly	
209					705					710					715		
211	cta	gat	gca	tca	cag	caa	atc	cgt	ctg	gaa	gca	gaa	gaa	gaa	aaa	tta	2208
212	Leu	Asp	Ala	Ser	Gln	Gln	Ile	Arg	Leu	Glu	Ala	Glu	Glu	Glu	Lys	Leu	
213	720					725					730					735	
215	gct	cat	ttg	cca	ctc	gct	aaa	aca	gta	tcc	gta	gaa	gag	ctt	ctg	caa	2256
216	Ala	His	Leu	Pro	Leu	Ala	Lys	Thr	Val	Ser	Val	Glu	Glu	Leu	Leu	Gln	
217					740					745						750	
219	tac	gtg	gag	ctt	caa	gat	cct	gtt	acg	cgc	acg	cag	ctt	cgc	gca	atg	2304
220	Tyr	Val	Glu	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr	Gln	Leu	Arg	Ala	Met	
221					755					760						765	
223	gct	gct	aaa	acg	gtc	tgc	ccg	ccg	cat	aaa	gta	gag	ctt	gaa	gcc	ttg	2352
224	Ala	Ala	Lys	Thr	Val	Cys	Pro	Pro	His	Lys	Val	Glu	Leu	Glu	Ala	Leu	
225					770					775						780	
227	ctt	gaa	aag	caa	gcc	tac	aaa	gaa	caa	gtg	ctg	gca	aaa	cgt	tta	aca	2400
228	Leu	Glu	Lys	Gln	Ala	Tyr	Lys	Glu	Gln	Val	Leu	Ala	Lys	Arg	Leu	Thr	
229					785					790						795	
231	atg	ctt	gaa	ctg	ctt	gaa	aaa	tac	ccg	gcg	tgt	gaa	atg	aaa	ttc	agc	2448
232	Met	Leu	Glu	Leu	Leu	Glu	Lys	Tyr	Pro	Ala	Cys	Glu	Met	Lys	Phe	Ser	
233	800					805					810					815	
235	gaa	ttt	atc	gcc	ctt	ctg	cca	agc	ata	cgc	ccg	cgc	tat	tac	tcg	att	2496
236	Glu	Phe	Ile	Ala	Leu	Leu	Pro	Ser	Ile	Arg	Pro	Arg	Tyr	Tyr	Ser	Ile	
237					820					825						830	
239	tct	tca	tca	cct	cgt	gtc	gat	gaa	aaa	caa	gca	agc	atc	acg	gtc	agc	2544
240	Ser	Ser	Ser	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala	Ser	Ile	Thr	Val	Ser	
241					835					840						845	
243	gtt	gtc	tca	gga	gaa	gcg	tgg	agc	gga	tat	gga	gaa	tat	aaa	gga	att	2592
244	Val	Val	Ser	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly	Glu	Tyr	Lys	Gly	Ile	
245					850					855						860	
247	gcg	tcg	aac	tat	ctt	gcc	gag	ctg	caa	gaa	gga	gat	acg	att	acg	tgc	2640
248	Ala	Ser	Asn	Tyr	Leu	Ala	Glu	Leu	Gln	Glu	Gly	Asp	Thr	Ile	Thr	Cys	
249					865					870						875	
251	ttt	att	tcc	aca	ccg	cag	tca	gaa	ttt	acg	ctg	cca	aaa	gac	cct	gaa	2688
252	Phe	Ile	Ser	Thr	Pro	Gln	Ser	Glu	Phe	Thr	Leu	Pro	Lys	Asp	Pro	Glu	
253	880					885					890					895	
255	acg	ccg	ctt	atc	atg	gtc	gga	ccg	gga	aca	ggc	gtc	gcg	ccg	ttt	aga	2736
256	Thr	Pro	Leu	Ile	Met	Val	Gly	Pro	Gly	Thr	Gly	Val	Ala	Pro	Phe	Arg	

## RAW SEQUENCE LISTING

DATE: 08/06/2002

PATENT APPLICATION: US/10/031,695

TIME: 10:35:30

Input Set : A:\031695s1

Output Set: N:\CRF3\08062002\J031695.raw

```

257          900          905          910
259 ggc ttt gtg cag gcg cgc aaa cag cta aaa gaa caa gga cag tca ctt 2784
260 Gly Phe Val Gln Ala Arg Lys Gln Leu Lys Glu Gln Gly Gln Ser Leu
261          915          920          925
263 gga gaa gca cat tta tac ttc ggc tgc cgt tca cct cat gaa gac tat 2832
264 Gly Glu Ala His Leu Tyr Phe Gly Cys Arg Ser Pro His Glu Asp Tyr
265          930          935          940
267 ctg tat caa gaa gag ctt gaa aac gcc caa agc gaa ggc atc att acg 2880
268 Leu Tyr Gln Glu Glu Leu Glu Asn Ala Gln Ser Glu Gly Ile Ile Thr
269          945          950          955
271 ctt cat acc gct ttt tct cgc atg cca aat cag ccg aaa aca tac gtt 2928
272 Leu His Thr Ala Phe Ser Arg Met Pro Asn Gln Pro Lys Thr Tyr Val
273 960          965          970          975
275 cag cac gta atg gaa caa gac ggc aag aaa ttg att gaa ctt ctt gat 2976
276 Gln His Val Met Glu Gln Asp Gly Lys Lys Leu Ile Glu Leu Leu Asp
277          980          985          990
279 caa gga gcg cac ttc tat att tgc gga gac gga agc caa atg gca cct 3024
280 Gln Gly Ala His Phe Tyr Ile Cys Gly Asp Gly Ser Gln Met Ala Pro
281          995          1000          1005
283 gcc gtt gaa gca acg ctt atg aaa agc tat gct gac gtt cac caa gtg 3072
284 Ala Val Glu Ala Thr Leu Met Lys Ser Tyr Ala Asp Val His Gln Val
285          1010          1015          1020
287 agt gaa gca gac gct cgc tta tgg ctg cag cag cta gaa gaa aaa ggc 3120
288 Ser Glu Ala Asp Ala Arg Leu Trp Leu Gln Gln Leu Glu Glu Lys Gly
289          1025          1030          1035
291 cga tac gca aaa gac gtg tgg gct ggg taa 3150
292 Arg Tyr Ala Lys Asp Val Trp Ala Gly
293 1040          1045
296 <210> SEQ ID NO: 2
297 <211> LENGTH: 1048
298 <212> TYPE: PRT
299 <213> ORGANISM: Bacillus megaterium
301 <400> SEQUENCE: 2
303 Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys Asn
304 1 5 10 15
306 Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys Ile
307 20 25 30
309 Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Arg Val
310 35 40 45
312 Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp Glu
313 50 55 60
315 Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Val Arg Asp
316 65 70 75 80
318 Phe Ala Gly Asp Gly Leu Phe Thr Ser Trp Thr His Glu Lys Asn Trp
319 85 90 95
321 Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala Met
322 100 105 110
324 Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val Gln
325 115 120 125

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/10/031,695

DATE: 08/06/2002  
TIME: 10:35:31

Input Set : A:\031695s1

Output Set: N:\CRF3\08062002\J031695.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:3; N Pos. 16,17,18  
Seq#:4; N Pos. 13,14,15  
Seq#:5; N Pos. 16,17,18  
Seq#:6; N Pos. 17,18,19  
Seq#:7; N Pos. 25,26,27  
Seq#:8; N Pos. 15,16,17  
Seq#:9; N Pos. 20,21,22  
Seq#:10; N Pos. 16,17,18  
Seq#:11; N Pos. 23,24,25  
Seq#:12; N Pos. 13,14,15  
Seq#:13; N Pos. 14,15,16  
Seq#:14; N Pos. 18,19,20  
Seq#:15; N Pos. 19,20,21  
Seq#:16; N Pos. 14,15,16  
Seq#:17; N Pos. 18,19,20  
Seq#:18; N Pos. 21,22,23  
Seq#:19; N Pos. 13,14,15  
Seq#:20; N Pos. 16,17,18